Voluntary Implementation

Matthew O. Jackson and Thomas R. Palfrey*

May 1998

This Revision: November 22, 1999

Abstract

We examine Nash implementation when individuals cannot be forced to accept the outcome of a mechanism. Two approaches are studied. The first approach is static where a state-contingent participation constraint defines an implicit mapping from rejected outcomes into outcomes that are individually rational. We call this voluntary implementation, and show that the constrained Walrasian correspondence is not voluntarily implementable. The second approach is dynamic where a mechanism is replayed if the outcome at any stage is vetoed by one of the agents. We call this stationary implementation, and show that if players discount the future in any way, then the constrained Walrasian correspondence is stationarily implementable.

*We thank an associate editor, an anonymous referee, and Bhaskar Dutta for helpful comments and suggestions. Division of Humanities and Social Sciences, Mail Code 228-77, Caltech, Pasadena, CA 91125

1 Introduction

Implementation theory has been successful in characterizing the objectives that a society can implement when accounting for the incentives that individuals have to take advantage of their information. Nevertheless, the theory is open to criticism for the sometimes implausible mechanisms (i.e., game forms) that it relies on to show which objectives may be achieved. In this paper we focus on remedying a specific, but critical, weakness of implementation theory: its use of implausible outcomes off the equilibrium path to enforce equilibrium behavior and/or to "break" undesirable equilibria (i.e., assure that undesired strategy combinations are not equilibria). The implausibility stems from the assumption that the outcome function is fully enforceable[1], which is not the case in many applications.

One source of difficulty in enforcement relates to commitment. If, for example, a mechanism is constructed to assist bargainers in reaching mutually improving agreements, then it is problematic to assume that highly inefficient outcomes will be allowed to stand. This is potentially a problem both on and off the equilibrium path, as off-equilibrium path considerations have implications for equilibrium behavior.

A second source of difficulty with enforcement relates to property rights that are exogenous to a mechanism and impose state-contingent constraints on a social choice rule. In many settings individuals have inalienable rights that guarantee them some outcomes in some states of the world. Many economic models treat these rights as exogenous, and only impose them as participation constraints or individual rationality constraints. Here, we stress the importance of considering these constraints out of equilibrium as well as in equilibrium.

A third source of difficulty with enforcement is related to dynamic contexts. Most implementation problems that have been studied are static. After the mechanism has reached an outcome, the world ends. In fact, most of these are more realistically viewed as being a single period of a multiperiod allocation problem. Thus, it makes sense to model explicitly the dynamics that can occur after the mechanism has tentatively reached an outcome. Is there renegotiation? Is the mechanism replayed? Is there time discounting between periods?

[1]See Hurwicz (1994) for a general discussion of issues related to enforceability in mechanism design.

The framework laid out is designed to address these problems with enforcement, and variations on them, in a unified way, and includes the standard implementation problem as a special case. While this admits a number of applications, the implications of the characterization theorems depend on the specifics of the setting. We then specialize to an application of the reasoning to a study of dynamic implementation in stationary equilibria, where specific implications can be understood.

1.1 Relation to the Literature

There has been a flurry of recent research into the general question of realistic restrictions on mechanisms.[2] There are several papers that address either the issue of individual rationality, or renegotiation. Most closely related to this paper are three papers that deal with imposing individual rationality or allowing for renegotiation both in and out of equilibrium.[3] Ma, Moore and Turnbull (1988) were the first to point out the importance of imposing an individual rationality constraint both in and out of equilibrium. They examined a principal-agent model where the usual individual rationality constraint (imposed only on the equilibrium path) was replaced by an "opt-out," where each player had the ability to decline the outcome of the mechanism and accept a status-quo outcome instead. Maskin and Moore (1998) examined a more general implementation problem, and changed the opting out to a possibility of renegotiation. They considered implementation where any outcome of a mechanism that suggests a Pareto dominated allocation is replaced by a Pareto efficient allocation according to an exogenous renegotiation function.[4] In Jackson and Palfrey (1998), in the context of a

[2] See Jackson (1997) for an overview and references.

[3] A more distantly related (but similarly motivated) problem in implementation theory is "credibility", or the ability of the planner to commit to off-equilibrium-path outcomes that are known to be undesirable, in order to implement desirable outcomes on the equilibrium path. Chakravorti, Corchon, and Wilkie (1992) investigate this, and Baliga, Corchon, and Sjöström (1995) and Baliga and Sjöström (1995) go further, by including the planner as a player in the mechanism.

[4] Rubinstein and Wolinsky (1992) took a different approach to incorporating the possibility of renegotiation into implementation. They examined "renegotiation-proof" implementation in a pairwise bargaining setting where the equilibrium was required to be immune to different sorts of renegotiation, and showed that the

dynamic bargaining and matching model, we endogenized the alternative coming from the "opt-out." We considered implementation when players have the ability to opt out of the outcome suggested by the mechanism and be rematched with a new bargaining partner. We showed that although such an endogenous individual rationality constraint is compatible with efficiency within individual matches, it could be incompatible with efficiency from society's point of view accounting for the overall evolution of a market.

Here, we begin by unifying these approaches. They all have the common feature of viewing a mechanism as an intermediate institution that suggests outcomes that may subsequently be altered. This may be captured in a general form of implementation where an outcome of a mechanism is converted by a general state-contingent allocation rule – which we call a reversion function. The characterization of implementable rules given such reversion functions follows a close parallel to the characterization of Nash implementable rules. Next, we examine voluntary implementation where the reversion function is in the form of an individual rationality constraint, in the spirit of Ma, Moore and Turnbull (1988), but taken to the general implementation problem. We show that the implications of such constraints may be derived in a variety of settings from voting to exchange. Finally, in the spirit of Jackson and Palfrey (1998), we examine a model where players may force the game form to be replayed, thereby endogenizing the reversion function (in this case, the alternative that individuals may be opting for). This fits well with the structure of many markets, where the mechanism represents the protocol or rules by which agents negotiate and trades are not finalized until all parties reach agreement. We show that without discounting, the set of implementable correspondences is severely limited, while with discounting much more positive results may be obtained, and for instance, the constrained Walrasian correspondence may be implemented.

The remainder of the paper is organized as follows. Section 2 lays out the general framework and explains how the standard implementation model is being extended. We establish necessary and sufficient conditions for voluntary implementation. The conditions are the natural extensions of monotonicity and no veto power, modified to incorporate the voluntary

possibilities for implementation depend on the way in which renegotiation is modeled.

constraints. Section 3 presents examples which illustrate the effect of voluntary constraints on the set of implementable social choice correspondences. That section also shows how individual rationality constraints, renegotiation, and blocking coalitions all fall within the bounds of this framework. Section 4 looks at implementation when the voluntary constraint is modeled as a replay of the mechanism.

2 Definitions

There is a finite set of *individuals* or agents, $I = \{1, \ldots, n\}$, a known set of *feasible outcomes*, denoted A, and a set of *states* S, with individual states denoted by s.

The *preferences* of individuals may be state dependent, and so each state has a corresponding profile of preference relations, $R(s) = (R_1(s), \ldots R_n(s))$, where $R_i(s)$ is a weak preference ordering over A. We write $aR_i(s)b$ if i weakly prefers a to b, and $aP_i(s)b$ if the preference is strict.

A *social choice correspondence* is a (possibly multi-valued) mapping $F : S \Rightarrow A$. A single-valued social choice correspondence is called a *social choice function*, and is denoted in the lower case, f. The set of all social choice correspondences is denoted by $\mathcal{F}$.

A *mechanism*, (M, g), consists of a *message space*, $M = M_1 \times \cdots \times M_n$ that is a Cartesian product of n individual message spaces, one for each agent, and an *outcome function*, $g : M \to A$.

2.1 Voluntary Implementation and h-Nash implementability

The idea behind voluntary implementation is similar to the notion of an individual rationality constraint or a participation constraint. Individuals are permitted to veto some subset of the feasible set, which may vary across states and individuals. This idea can be illustrated in the simple example of a pure exchange economy with fixed initial endowments, $\omega = (\omega_1, \ldots, \omega_n)$. If x denotes some reallocation of ω, then this reallocation is individually rational at s if and only if $x_i R_i(s) \omega_i$ for all i. Suppose the mechanism is (M, g), the players report m at s, and the reallocation specified by the outcome function is $g(m)$. If $\omega_i P_i(s) g_i(m)$ for some i, and we wish the mechanism to reflect voluntary trade, then we should allow this individual to veto the outcome $g_i(m)$. The issue, of course, is how to specify the consequences of this veto.

The simplest way to model voluntary implementation is to explicitly specify what happens if an individual vetoes an outcome. To do this involves specifying a function that maps states into allocations.

A *reversion function,* $h : S \to A$, is a mapping that indicates what the outcome is in the case of a veto by some individual. A reversion function h induces a mapping $H : A\times\ S \times \mathcal{F}$, by

$$\begin{aligned} H(a,s,h) &= a \quad \text{if } a\ R_i(s)\ h(s) \text{ for all } i \\ &= h(s) \quad \text{otherwise.} \end{aligned}$$

An action profile m is an *h-Nash equilibrium* of (M, g) at s if

$$H(g(m),s,h)\ R_i(s)\ H(g(\widehat{m}_i,m_{-i}),s,h) \ \text{ for all } i\ \widehat{m}_i \in M_i$$

A social choice correspondence F is *h-Nash implementable* if there exists a mechanism, (M, g) such that, for all s :

(i) For each $a \in F(s)$ there exists an h-Nash equilibrium, $m \in M$, such that $H(g(m),s,h) = a$

(ii) If $m \in M$ is an h-Nash equilibrium at s, then $H(g(m),s,h) \in F(s)$.

2.2 Necessary condition for h-Nash implementation

It is well-known that monotonicity of the social choice correspondence is a necessary condition for Nash implementation. (See Maskin (1998))

F is *monotonic* if, for all $s, s' \in S$ and $a \in F(s)$ such that $a \notin F(s')$, there exists $b \in A$ and $i \in I$ such that $a\ R_i(s)\ b$ and $b\ P_i(s')\ a$.

The intuition behind this condition is that if $a \in F(s)$ but $a \notin F(s)$, then implementability of F implies the existence of a mechanism where a is a Nash equilibrium outcome at s, but not a Nash equilibrium outcome at s'. Thus, considering the equilibrium strategies leading to a at s, there must exist an agent i that has a deviation (resulting in b), which must be preferred by i at s', but not at s.

This condition generalizes in a straightforward way to h⇔implementation. We call this condition reversion-monotonicity.

A social choice correspondence, F, is *reversion-monotonic* relative to h if, for all $s \in S$ and for all $a \in F(s)$, there exists $z \in A$ such that:

1. $H(z, s, h) = a$, and
2. For all $s' \in S$ such that $H(z, s', h) \notin F(s')$, there exists $y \in A$ and $i \in I$ such that $H(z, s, h)\ R_i(s)\ H(y, s, h)$ and $H(y, s', h)\ P_i(s')\ H(z, s', h)$.

The necessity of this condition follows the same reasoning as the necessity of monotonicity for Nash implementation. There are two differences however. The first is noted in item 1 above, where it is recognized that a may not be coming directly from the mechanism, but instead from the reversion function. The second difference is in item 2, where it is not just the lower contour set of a that matters, but also the (state-dependent) reversion function, since this function determines what outcomes will be vetoed in each state and the resulting reversion point following a veto.

2.3 Generalized Reversion Functions

In fact, this necessary condition can be stated in a more general form, which will prove useful in the dynamic context as well. The approach outlined above with a reversion function presumes that any single agent can veto an outcome and then the alternative that replaces it is independent of the starting alternative. Instead, we can consider the situation where any suggested alternative a is converted in a state dependent way via some mapping G.

Consider any mapping $G : A \times S \to A$. The H defined above for a given h is one such function.

We say that m is an *G-Nash equilibrium* of (M, g) at s if

$$G(g(m), s)\ R_i(s)\ G(g(\widehat{m}_i, m_{-i}), s) \text{ for all } i\ ,\ \widehat{m}_i \in M_i$$

A social choice correspondence F is *G-Nash implementable* if there exists a mechanism, (M, g) such that, for all s :

(i) For each $a \in F(s)$ there exists an G-Nash equilibrium, $m \in M$, such that $G(g(m), s) = a$

(ii) If $m \in M$ is a G-Nash equilibrium at s, then $G(g(m), s) \in F(s)$.

A social choice correspondence, F, is *G-monotonic* if, for all $s \in S$ and for all $a \in F(s)$, there exists $z \in A$ such that:

1. $G(z, s) = a$

2. For all $s' \in S$ such that $G(z, s') \notin F(s')$, there exists $y \in A$ and $i \in I$ such that $G(z, s)\ R_i(s)\ G(y, s)$ and $G(y, s')\ P_i(s')\ G(z, s')$.

The following theorem follows directly from the logic of Maskin's theorem (1998).

Theorem 1 *If F is G-Nash implementable, then F is G-monotonic.*

Proof: Consider a state, s, and an outcome, $a \in F(s)$. Let (M, g) G-implement F in Nash equilibrium, and let m be a Nash equilibrium at s which produces a as the outcome. That is, $G(g(m), s) = a$. Next suppose that $G(g(m), s') \notin F(s')$ for some other state s'. Let $z = g(m)$. Since m is a G-Nash equilibrium at s, it must be that $G(g(m), s)\ R_i(s)\ G(g(\widehat{m}_i, m_{-i}), s)$ for all i, $\widehat{m}_i$. But since F is G-Nash implementable and $G(z, s') \notin F(s')$, we know that m is not a G-Nash equilibrium at s'. So, there exists i and $\widehat{m}_i$) such that $G(g(\widehat{m}_i, m_{-i}), s')P_i(s')(g(m), s')$. Let $y = g(\widehat{m}_i, m_{-i})$, to satisfy the definition of G-monotonicity. ∎

Similarly, sufficient conditions for Nash implementation have analogs for G-Nash implmentation.

With Nash implementation, it is well-known that if there are at least 3 players, monotonic social choice correspondences are Nash implementable if they satisfy No Veto Power. No veto power states that if all players except possibly one agree on a best outcome in some state, then that outcome must be in the social choice correspondence at that state. A similar result follows here for G-Nash implementation, using an appropriately modified version of NVP.

A social choice correspondence F satisfies *G-No Veto Power (G-NVP)* if, for all $i \in I$, for all $j \neq i$, $z \in A$, and $s \in S$, then $G(z,s) \in F(s)$ whenever $G(z,s)R_j(s)G(y,s)$ for all $y \in A$.

Theorem 2 *If $n \geq 3$ and F is G-monotonic and satisfies $G\Leftrightarrow NVP$, then F is G-implementable.*

Again, the proof is an easy extension of proofs of Nash implementability, and is provided in an appendix.

3 Applications

3.1 Implementation with Individual Rationality Constraints

One of the most natural applications of h-implementation is to problems in which there is a fixed status quo outcome that any agent can revert to. For example, in the case of exchange economies, it is often natural to assume that each individual can protect their initial endowment. Surprisingly, applications of implementation theory to exchange environments generally ignore these constraints. This is not to claim that implementation theory has not investigated whether certain individually rational social choice functions are implementable. That is a much different issue. The issue is that individual rationality constraints must be respected for the entire outcome function, rather than just at the equilibrium outcome. Why? If the individual messages happen to produce an outcome that violates individual rationality constraints, then a violated agent can simply veto the outcome. We are requiring that the mechanism be voluntary: no agent can be forced to accept an outcome. This views the mechanism as the protocol for communication and negotiation between agents, after which all of their signatures are required before a suggested outcome becomes final.[5]

As we will see below, the constraint that outcomes be acceptable to all agents can either restrict or even expand the set of allocation rules that are implementable. The intuition for why the set of implementable social choice functions can be restricted by such constraints is obvious. The intuition for why the set of implementable social choice functions can expand

[5]Thus, this viewpoint takes a mechanism as the means by which binding contracts are formed, rather than viewing the mechanism as a binding contract itself.

is more subtle, and has to do with the fact that these constraints implicitly provide state-contingent threat points that can affect equilibrium behavior.[6]

The simplest reversion function is simply a fixed status quo outcome, w, which results if any individual vetoes $g(m)$. That is, $h(s) = w$ for all s and

$$\begin{aligned} H(a, s, h) &= a \text{ if } aR_i(s)w \text{ for all } i \\ &= w \text{ if } wP_i(s)a \text{ for some } i. \end{aligned}$$

We call h-implementation with this kind of reversion function *IR-implementation*, and we refer to h-reversion monotonicity with this kind of reversion function as *IR-monotonicity.*

The following examples illustrate how IR-monotonicity can differ from monotonicity. The first example illustrates this surprising phenomenon that a social choice correspondence may satisfy IR-monotonicity, but fail to be monotonic.

Example 1 (Voting)

Let $A = \{w, x, y, z\}$, $I = \{1, 2, 3\}$, and $S = \{s, s'\}$. The status quo outcome is w (regardless of the state). Preferences are described below, where higher outcomes in the table are preferred to lower outcomes.

	s			s'	
1	2	3	1	2	3
x	y	z	x	y	z
z	z	x	z	z	x
y	x	w	y	x	y
w	w	y	w	w	w

Let $F(s) = \{x, z\}$ and $F(s') = \{x\}$. F is not monotonic, since $z \in F(s)$, $z \notin F(s')$, but the only preference reversal between s and s' involves agent 3's preferences changing between outcomes y and w. However, F satisfies IR-monotonicity. To see this, note that $H(y, s', h) = y$,

[6] The fact that allowing agents to opt-out can ease implementation has been previously noted in a moral hazard setting by Arya, Glover, and Hughes (1997).

but $H(y, s, h) = w$. Thus, for player 2 it is the case that $H(z, s, h)R_2(s)H(y, s, h)$ and $H(y, s', h)P_2(s')H(z, s', h)$, since these two relations reduce to $zR_2(s)w$ and $yP_2(s')z$, respectively.

To understand this phenomenon, note that the revision function introduces a form of sequential rationality to the Nash implementation problem.

The same phenomenon can be seen in an exchange economy.

Example 2 (An exchange economy)

Consider a two-person two-good exchange economy, with initial endowment point $w = ((1, 5), (5, 1))$. There are two states, which determine two possible preference profiles. In state s, both players have preferences represented by symmetric Cobb-Douglas utility functions $U(x_{i1}, x_{i2}) = x_{i1}x_{i2}$. In state s' individuals have Leontief preferences represented by $U(x_{i1}, x_{i2}) = \min\{x_{i1}, x_{i2}\}$. This is shown in figure 1.

FIGURE 1 HERE

Consider the following selection from the Pareto correspondence:[7]

$$\begin{aligned} f(s) &= x^{CD} = ((3, 3), (3, 3)) \\ f(s') &= x^{L} = ((2, 2), (4, 4)) \end{aligned}$$

The social choice function f is not monotonic. To see this, note that monotonicity requires that if $x \in f(s)$ and $x \notin f(s')$, then there exists i and y such that $yP_i(s')x$ and $xR_i(s)y$. In this case, given the Leontief preferences at s', if $yP_i(s')x^{CD}$ then $y_i \geq (3, 3)$ and $y_i \neq (3, 3)$. But this implies that $yP_i(s)x^{CD}$ given the Cobb-Douglass preferences at s. However, while f is not monotonic, it is in fact IR-monotonic. The key is that the lower contour sets relative to w differ between states s and s'. To see how the problem above with monotonicity is

[7]A similar example appears in Moore and Repullo (1988), to illustrate how non-monotonic social choice functions can be implemented in subgame perfect equilibrium.

overcome, note that IR-monotonicity requires that if $x \in f(s)$ then there exists z such that $x = H(z,s,h)$ (where h reverts to the endowment w) and if $H(z,s,h) \notin f(s')$ then there exists y and i such that $H(z,s,h)\ R_i(s)\ H(y,s,h)$ and $H(y,s',h)\ P_i(s')\ H(z,s',h)$. Here, let $z = x^{CD}$, $y = x^L$ and $i = 2$. Then $H(y,s,h) = w$ as agent 1 vetoes x^L in state s while $H(y,s',h) = x^L$. So the condition is satisfied as we have $x^{CD}R_i(s)w$ and $x^L P_i(s')x^{CD}$.

Moreover, f is IR-implementable via the trivial mechanism where player 2 simply chooses between x^{CD} and x^L. In state s, x^{CD} is individually rational, but x^L is not individually rational for player 1, so this choice reduces to a choice between x^{CD} and w. Since player 2 prefers x^{CD} to w, his optimal choice is x^{CD}. In state s', both x^{CD} and x^L are individually rational for both players. Since player 2 prefers x^L to x^{CD}, his optimal choice is x^L. Thus, this simple mechanism voluntarily implements the stated allocation rule, which is not Nash implementable.

Both examples 1 and 2 show that there are voluntarily implementable social choice correspondences that are not Nash implementable. The next example shows the converse.

Example 3 (Nash Implementable but not Voluntarily Implementable)

Let $A = \{w,x,y\}$, where w is the status quo. Let $I = \{1,2\}$. Let $S = \{s,s'\}$. Preferences are described by:

s			s'	
1	2		1	2
x	x		w	y
y	y		y	x
w	w		x	w

$$\begin{aligned} F(s) &= \{x\} \\ F(s') &= \{y\} \end{aligned}$$

This social choice function is monotonic and implementable by the simple mechanism where player 2 chooses between x and y. However, this is not individually rational in state

s', and hence not voluntarily implementable. In particular, individual rationality requires $F(s') = \{w\}$.

Next we show that the constrained Walrasian correspondence is an important social choice correspondence that falls into the category of being Nash implementable, but failing to be voluntarily implementable.

Example 4 (Non-Implementability of the Constrained Walrasian Correspondence)

Consider a two-person two-good exchange economy, with initial endowment point w. There are two states, which determine two possible preference profiles, as illustrated in figure 2 below.

FIGURE 2 HERE

Here, the unique Walrasian outcome at s is not a Walrasian equilibrium at s'. However, the only changes in preferences relative to a occur at points that are not individually rational for agent 2. Since such points will always be vetoed and lead to w, there are no preference reversals between s and s' that can be used to satisfy IR-monotonicity. Thus, any mechanism that yields a as an IR equilibrium outcome at s must also produce a as an IR equilibrium outcome at s', even though it is not a Walrasian outcome at s'. Since these are interior points, this applies to the constrained Walrasian correspondence.[8]

The next example shows that the implications of voluntary implementation extend far beyond the consideration of Nash implementation. There are also implications for other forms of implementation.

Example 5 (Not Voluntarily Implementable via any Solution, but Implementable via Many)

[8]See Hurwicz, Maskin and Postlewaite (1995) for a detailed discussion of the constrained Walrasian correspondence and its Nash implementability.

This is an example of an allocation rule that is individually rational, and is implementable in subgame perfect equilibrium, undominated Nash equilibrium, iterative elimination of weakly dominated strategies, and is virtually implementable. However, it is not implementable by any solution concept if agents can veto outcomes that are not individually rational. As in example 2, denote the initial endowment point by w. There are two states, and the utility functions in state s are called U_1^S and U_2^S, respectively. In state s', $U_1^{S'}$ is identical to U_1^S for all allocations x for which $U_1^S(x_1) \geq U_1^S(w_1)$ and $U_2^{S'}$ is identical to U_2^S for all allocations x for which $U_2^S(x_2) \geq U_2^S(w_2)$. That is, preferences differ only on allocations outside the set of individually rational allocations. Let $f(s) = x$ and $f(s') = x'$ be an individually rational allocation rule, shown in figure 3.

FIGURE 3 HERE

Let U_1^S and $U_1^{S'}$ differ outside of the individually rational lens, as shown in that same figure. This social choice function violates IR-monotonicity, since the utility functions only differ on allocations that will revert to w in any mechanism that specifies them in the outcome function.

In constrast, it is easy to show that these allocation rules are implementable via subgame perfect equilibrium, undominated Nash equilibrium, iterated weak dominance, perfect equilibrium, and is also virtually implementable. For example, the following mechanism implements f via iterated weak dominance, where y and z are the allocations marked in figure 2:

	m_{21}	m_{22}
m_{11}	x	y
m_{12}	x	z
m_{13}	x'	x'

In state s, m_{11} weakly dominates m_{12}. At the next iteration, m_{22} is weakly dominated by m_{21}. At the third and last iteration m_{11} strictly dominates m_{13} so the solution in state s

is m_{11}. In state s', m_1 weakly dominates m_{11}. At the next iteration, m_{22} weakly dominates m_{23}. At the third and last iteration m_{13} strictly dominates m_{12} so the solution in state s' is m_{13}. Similar mechanisms can be constructed for implementation by other refinements.

The insight from this example is that the voluntary constraint implies that individuals' preference relations over outcomes that are not individually rational (for some individual) are irrelevant. Refinements have been used in implementation theory to take advantage of any reversal in preferences, even when these involve alternatives that are sub-optimal or not individually rational. This is not possible in voluntary implementation, regardless of the solution concept used.

3.2 Implementation with Renegotiation

Maskin and Moore [1998] consider a different version of a reversion function, which also fits nicely within the present framework. They are concerned with the renegotiation problem that can arise in mechanism design. In particular, they argue that if $g(m)$ is inefficient in state s, then the players will renegotiate the outcome to something that is Pareto efficient, and which Pareto dominates it. Since the second property (Pareto domination of $g(m)$) will generally depend on $g(m)$ itself, they define a reversion function that depends not only on the state but also on the the allocation that is vetoed. In particular, they define a renegotiation function $r : A \times S \to A$. Given that r is Pareto efficient, there will always be some voter who would veto $g(m)$ if it were inefficient at stage s. Therefore, implementation with renegotiation is consistent with our "veto" interpretation of the h function, and is an example of a G function.

3.3 Implementation with Coalitional Veto Sets

The notion of h-implementation can be generalized substantially, within the framework of the G-function. First, as with implementation with renegotiation, one can allow h to depend on the outcome that is being vetoed.[9] Second, and perhaps more interesting, one can allow for

[9]Implicitly, the G function can also incorporate information about who is vetoing, as G depends on the state. This then allows for outcomes such as veto by one agent and trade amongst the remaining agents.

coalitional veto sets. For example, one can require majority rule approval of the outcome of the mechanism, with the outcome reverting to $h(s)$ if $g(m)$ is does not receive a majority. The generalization of this is the concept of blocking coalitions. In our definition of "voluntary" implementation, each individual constitutes a blocking coalition. In many contexts, one can argue that this is too strong a requirement, and that larger coalitions may be needed to veto an outcome.

Let $C : S \Rightarrow 2^I$ be a blocking coalition correspondence, which specifies the set of all blocking coalitions in each state. Thus, for example, under voluntary implementation, $C(s) = 2^I \Leftrightarrow \emptyset$ for all s, so that any set of objecting individuals can prevent an outcome. This defines the mapping, $H^C : A \times S \times \mathcal{F}$, by

$$\begin{aligned} H^C(a,s,h) &= a \text{ if, for all } c \in C(s), \quad aR_i(s)h(s) \text{ for some } i \in c \\ &= h(s) \text{ otherwise} \end{aligned}$$

This fits the form of a G-function and so the results on G-implementation apply.

While the examples and applications described above show the broad coverage of general reversion function techniques for analyzing implementation, much of the details of implementability depend on the specification of the reversion function. Rather than simply take that function as being exogenous, we turn next to analyze a class of situations where the reversion function is naturally endogenously determined.

4 Voluntary Implementation with Repeated Mechanisms

In this section, we endogenize the generalized reversion function, by considering situations where a player opting out of an outcome simply forces the mechanism to be replayed. The motivation for examining this situation is simple, and related to the motivation for studying implementation with renegotiation. If an individual vetoes $g(m)$, it is unnatural to suppose that the world stops at that moment. For example, in a pure exchange environment, if an agent vetoes $g(m)$, and the endowment results, the individuals in the economy could simply play the mechanism again.[10] This captures applications where we think of the mechanism as describing the methods or framework available to agents for communication and negotiation, where the replay of the mechanism is the natural form of (re)-negotiation.

This is how game theorists have modeled bargaining. When two agents bargain, say by offers and counteroffers, rejection of an offer generally does not mean no-trade (except in very special cases like the ultimatum game). The reason for this is that the notion of *voluntary* trade implies that if there are still gains to trade to be exploited, the agents involved will continue playing some game. In this section, we explore a general model of recontracting of this sort, when rejection (i.e., veto) of $g(m)$ is followed by simply replaying the mechanism again in the following period. This converts the original mechanism into an infinite game form.

For this reason, we time date outcomes, so the outcome space is expanded to be $\overline{A} = A \times \{1, 2, 3 \ldots\}$ and a typical outcome is denoted a_t. For simplicity we write drop the subscript in the first period and write $a_1 = a$. In the event that players use strategies such that no outcome is ever reached, the outcome of the game is denoted $\emptyset$. We assume that $y_t P_i(s) \emptyset$ for all i, y, s, and t.[11]

4.1 Stationary Preferences, Equilibrium, and Implementation

Preferences are extended to be complete and transitive on $A \times \{1, 2, 3 \ldots\}$ for all players. The following assumptions on extended preferences over time dated outcomes, capture a stationarity of preferences.

1. $a_t R^i(s) b_t \Leftrightarrow a_{\tilde{t}} R^i(s) b_{\tilde{t}}$ for all $a, b \in A$ and $t, \tilde{t} \in \{1, 2, 3 \ldots\}$ (*Ordinal stationarity*).
2. $a_t R^i(s) b_{t+1} \Leftrightarrow a_{\tilde{t}} R^i(s) b_{\tilde{t}+1}$ for all $a, b \in A$ and $t, \tilde{t} \in \{1, 2, 3 \ldots\}$ (*Intertemporal stationarity*)

[10] More generally, a veto might trigger an alternative mechanism which is played. We looked at voluntary implementation using sequential mechanisms in a matching/bargaining framework in Jackson-Palfrey (1998).

[11] This is a simplifying assumption. It only needs to be true that implemented outcomes are weakly preferred to no outcome, and that there is some outcome at some date that is strictly preferred by all agents to no outcome. See Lemma 1 in the appendix and its proof for details.

3. $a_t R^i(s) a_{t+1}$, for all i, a, s, and $t = 1, 2, \ldots$. (*Weak Impatience*).

The above assumptions are maintained throughout this section.

The first two restrictions on preferences guarantee that individuals' tastes do not change over time, and are thus time consistent. The third restriction avoids the pathological case where individuals always prefer to defer agreement to the future.

A message profile m is a *stationary equilibrium* of (M, g) at s if, for all i, and $\widehat{m}_i \in M_i$, $g(m)\ R_i(s)\ H(g(\widehat{m}_i, m_{-i}), s, g(m)_2)$.[12]

So, a message profile is a stationary equilibrium if each player is best responding knowing that a veto today results in the same message profile being played tomorrow. Essentially, stationary equilibria correspond to the Markov perfect equilibria of the game form where in a given period the mechanism is played, then agents are called on to veto sequentially, and the process terminates with $g(m)$ if there is no veto and starts over in the next period if there is a veto. To be precise, in the appendix we show that the set of stationary equilibria correspond exactly to the set of Markov perfect equilibria of the game form described above, where agents do not veto when indifferent.

While the stationary equilibria have a foundation as Markov perfect equilibria and may be argued for on the grounds of simplicity, restricting attention to such equilibria rules out some behavior that may be quite natural. Most importantly, such strategies ignore history and eliminate many folk-theorem like constructions.[13]

A social choice function is **attainable** in stationary equilibrium via a mechanism (M, g) if, for each s, there exists a stationary equilibrium m_s such that $g(m_s) = f(s)$.

Attainability is a very weak form of implementation (essentially, an indirect version of truthful implementation). A social choice correspondence F is **implementable in stationary equilibrium** if there exists a mechanism, (M, g) such that, for all s :

[12]Note that here the last argument of H is an outcome rather than a social choice function. This obvious extension can be made formal by considering the constant social choice function resulting in $g(m)_2$.

[13]See Baron and Ferejohn (1989) for examples of such constructions and the role of stationarity in multilateral bargaining.

(i) For each $a \in F(s)$ there exists a stationary equilibrium, $m \in M$, such that $g(m) = a$

(ii) If $m \in M$ is a stationary equilibrium at s, then $g(m) \in F(s)$.

Before taking a careful look at stationary implementation, we first apply Theorems 1 and 2 to characterize stationary implementation. Given individuals' weak preference against delay, the definition of G-monotonicity translates to:

A social choice correspondence F satisfies **stationary monotonicity** if, for all s, s', and for all x such that $x \in F(s)$ but $x \notin F(s')$, there exists $y \in A$ and $i \in I$ such that for all t, $x_t\ R_i(s)\ H(y_t, s, x_{t+1})$ and $H(y_t, s', x_{t+1})\ P_i(s')\ x_t$.

Theorem 3 *If a social choice correspondence is implementable in stationary equilibrium then it satisfies stationary monotonicity.*

Proof: This follows from Theorem 1. ∎

Within this abstract framework, we can also obtain a standard characterization of sufficiency for the case of 3 or more agents. If a social choice function satisfies stationary monotonicity and an appropriately modified version of NVP, then it is implementable in stationary Nash equilibrium. The modification of NVP to the dynamic case is stated below.

A social choice correspondence F satisfies **stationary No Veto Power** if for any i, $z \in A$, and $s \in S$

$$[z_t R_j(s) H(y, s, z_{t+1})\ \forall y \in A, t, j \neq i] \Rightarrow [z \in F(s)].$$

Theorem 4 *If $n > 2$ and a social choice correspondence satisfies stationary monotonicity and stationary NVP, then it is implementable in stationary equilibrium.*

Proof: This follows from Theorem 2. ∎

Let us now analyze environments with more structure, where we can get a more detailed picture of stationary implementation.

4.2 Stationary Implementation with No Discounting

First, we consider the case with no discounting. That is the case where $a_t\ I^i(s)\ a$, for all i, a, s, and $t = 1, 2, \ldots$. Therefore, the time subscripts in the definition of stationary equilibrium can be removed, and the problem becomes very straightforward.

In this case, we call f *self-attainable* if it is attainable in stationary equilibrium.

Self-attainability simply considers which social choice functions can be supported as stationary equilibria of a mechanism, completely ignoring the multiple equilibrium problem usually at the heart of implementation theory. The next proposition shows a simple sufficient condition for self-attainability.

Let PE denote the Pareto correspondence

$$PE(s) = \{x \ |\forall y, yP_i(s)x \Rightarrow \exists j \text{ s.t. } xP_j(s)y\}.$$

We say that a social choice function f is Pareto efficient if $f(s) \in PE(s)$ for every s.

Proposition 1 *If f is Pareto efficient, then f is self-attainable.*

Proof: Consider the mechanism in which each player simultaneously announces an outcome, so $M_i = A$. If all announcements match, say $m_i = x$ for all i, then let $g(m) = x$. If the announcements don't all match, then let $g(m) = x_0$, where x_0 is some pre-specified default outcome. Consider s, and $x = f(s)$. Note that x is Pareto efficient at s. We need only show that $m_i = x$ for all i and t forms a stationary equilibrium at s. Suppose to the contrary that there exists i such that $H(g(\widehat{m}_i, m_{-i}), s, g(m)_2)P_i(s)\ g(m)$. This implies that $H(g(\widehat{m}_i, m_{-i}), s, g(m)_2) = g(\widehat{m}_i, m_{-i})$. However, by the Pareto efficiency of $g(m)$, it follows that since $g(\widehat{m}_i, m_{-i})P_i(s)\ g(m)$, there must exist j such that $g(m)P_j(s)g(\widehat{m}_i, m_{-i})$. This contradicts the fact that $H(g(\widehat{m}_i, m_{-i}), s, g(m)_2) = g(\widehat{m}_i, m_{-i})$. ∎

While any Pareto efficient selection is self-attainable, it is clear that the mechanism outlined in the above proof has a multitude of equilibria, some of which can be inefficient. So we should be interested in understanding which social choice correspondences are implementable in this setting. So, we turn to the stronger notion of implementation in stationary equilibrium, which we call *self-implementation* in the no discounting case.

Let R_F denote the range of F.

Proposition 2 *If F is self-implementable, then $R_F \cap PE(s) \subset F(s)$.*

Proof: Consider s and $a \in R_F \cap PE(s)$, and a mechanism (M, g) that self-implements F. Since $a \in R_F$ it follows that there exists m such that $g(m) = a$. Since a is Pareto efficient, then m is a self-equilibrium at s by an argument similar to that in the proof of Proposition 1. ∎

In special cases, we can say more. For example, if agents all have strict preferences over the Pareto set in any state, then the Pareto Correspondence is self implementable. Consider a mechanism such that for every profile of actions of the other agents, each agent has an action which can lead to any outcome. (Such a mechanism exists, as evidenced by a modulo construction.) It is easily seen that any such mechanism fully self-implements the Pareto correspondence.

Another implication of Proposition 2 is that the (constrained) Walrasian correspondence is not self-implementable, without discounting. This can be seen by revisiting Example 4, where a is in the range of the (constrained) Walrasian correspondence and a is Pareto efficient at s', but a is not a (constrained) Walrasian equilibrium at s'.[14]

Further implications of Proposition 2 depend on the structure of R_F. To see this, consider two extremes. At one extreme $R_F = \{y\}$, so F is constant and $F(s) = \{y\}$ for every s. This is obviously self-implementable by any trival mechanism that has $g(m) = y$ for all m. Thus one can self-implement very selective F's that may not be Pareto efficient. At the other extreme suppose that $R_{PE} \subset R_F$. Thus, the range of F is quite large and includes all allocations that may be Pareto efficient at some s. In this case, Proposition 2 implies that $PE(s) \subset F(s)$. This means that of the range of F is sufficiently rich, then F cannot be selective among Pareto efficient allocations but must include them all.

[14] Note that while the partial linearity of the indifference curve through a and w was necessary for Example 4, it is not necessary for this point, and one can easily find similar examples with strictly convex preferences.

Thus Proposition 2 shows that the apparent permissiveness of proposition 1 is deceiving. On the one hand it is true that any Pareto efficient allocation rule self-attainable. On the other hand, once one constructs a mechanism to attain that allocation rule, then any other allocation rule that is efficient relative to the range of the mechanism is also a stationary equilibrium outcome of the mechanism. Finer selections from the Pareto efficient set of allocation rules (relative to the range of the mechanism) are not self-implementable. So, effectively one can only be selective at a given s by making sure that the range of F is narrow across all s.

Note that these propositions are reminiscent of findings in bargaining theory (e.g., Rubinstein (1982)). For example, if individuals do not discount the future, then any bargaining split is an equilibrium outcome of the Rubinstein-Stahl bargaining game. What we have shown above is that this feature is robust across settings and mechanisms. We also know from the bargaining literature, that the introducing a strict preference against delay changes the scope of equilibrium. That turns out to be true in the dynamic implementation problem as well, as we now explore.

4.3 Stationary Implementation with Discounting

In most settings, it is more reasonable to expect that rejection of the outcome of a mechanism will lead to delay in the implementation of a final outcome, and that individuals find this delay costly.

The discounting case is formalized by requiring that $aP_i(s)a_t$, for all i, a, s, and $t > 1$.

For the discussion of the discounting case we assume that A is a metric space, with metric $|\cdot|$.

To begin to understand stationary implementation with discounting, we study a useful strengthening of stationary monotonicity. This strong form of monotonicity is applicable in many settings, including exchange economies.

A social choice correspondence F satisfies **local monotonicity** if, for all s, s', and $x \in F(s)$ such that $x \notin F(s')$, and for all ε, there exists $y \in A$ and $i \in I$ with $|y \Leftrightarrow f(s)| < \varepsilon$ such that $x \ R^i(s) \ y$ and $y \ P^i(s') \ x$.

Local monotonicity is a strengthening of monotonicity in that it requires that the test alternative y can be picked to be arbitrarily close to x. If preferences are continuous then y can be chosen so that it is preferred to receiving $f(s)$ with one-period delay. This leads to the following theorem that shows that in some reasonable cases, local monotonicity implies stationary monotonicity.

Theorem 5 *If preferences are continuous and individuals discount the future, then F satisfies local monotonicity only if it satisfies stationary monotonicity.*

Proof: Consider s, s' and $x \in F(s)$ such that $x \notin F(s')$. By the continuity of preferences and discounting, there exists ε such that $zP_j(s')x_2$ for all j and z such that $|z \Leftrightarrow x| < \varepsilon$. Apply local monotonicity with this ε, to find i and y with the properties stated in the definition of local monotonicity. By our choice of ε it follows that $x_t R_i(s) H(y_t, s, x_{t+1})$ and $H(y_t, s', x_{t+1}) = y_t P_i(s') x_{t+1}$. Thus, stationary monotonicity is satisfied. ∎

We now illustrate the power of Theorem 5, by applying it to pure exchange environments to show that the constrained Walrasian social choice function can be implemented in stationary equilibria.

Let ℓ denote the number of goods and $e_i \in \mathbb{R}^\ell_+$ denote the endowment of agent i, where $\sum_i e_i \in \mathbb{R}^\ell_{++}$. Here, $A = \{x \in \mathbb{R}^{n\ell}_+ | \sum_i x_i \leq \sum_i e_i\}$. For each i, the preferences of i depend only on i's allocation (so $xI_i(s)y$ whenever $x_i = y_i$), are continuous, increasing (where $x_i \geq y_i$ implies $xP_i(s)y$), and convex.

An allocation $x \in A$ is a *constrained Walrasian equilibrium* at s if there exists $p \in \mathbb{R}^\ell_+$ such that

- $p \cdot x_i \leq p \cdot e_i$
- $xR_i(s)y$ for all $y \in A$ such that $p \cdot y_i \leq p \cdot e_i$.

The constraint in the definition above appears in the restriction to $y \in A$ which implies that i's demand of any good is limited by the total available endowment in the economy.

Theorem 6 *With time discounting preferences, if $n \geq 3$, then the constrained Walrasian correspondence is implementable in stationary equilibrium.*

Proof: We need only show that the constrained Walrasian correspondence satisfies local monotonicity. Theorem 6 then follows from Theorems 5 and 4.[15] Consider x that is a constrained Walrasian equilibrium at s, with corresponding price p, but not at s'. It follows that x is not the (constrained) demand of some agent i at s'. Given that preferences are continuous, convex, and increasing, it follows that for any ε, there exists $y \in A$ such that $p \cdot y_i \leq p \cdot x_i$ and $|y \Leftrightarrow x| < \varepsilon$ and $y_i P_i(s')x_i$. Local monotonicity is thus satisfied. ∎

It is essential to Theorem 6 that the set of alternatives not be discrete. The ability to find trade-offs locally is critical to the theorem. This is analagous to results in the theory of bargaining, as for instance in the game analyzed by Rubinstein (1982) if only discrete offers can be made, then if agents do not discount too much, then all splits are sustainable in stationary equilibrium. Here, similar results hold: the results of the no-discounting case carry over if the set of alternatives is discrete and players are sufficiently patient.

5 Concluding Remarks

We have analyzed an approach to implementation with generalized reversion functions, that unifies and extends some analyses of agents' abilities to opt out of a mechanism, renegotiate, or otherwise alter the suggested outcome of a mechanism. We have also showed how this (static) implementation approach can be usefully applied to dynamic settings to understand implementation via stationary equilibria, where agents may veto a tentative outcome of a mechanism and opt instead to play the mechanism over again.

There is a rich array of applications where dynamics are a crucial element, ranging from the operation of continuous trading institutions to the rules governing electoral and legislative institutions. While the analysis here (see also Kalai and Ledyard (1998)) gives us some initial

[15] It is straightforward to check that stationary NVP is satisfied in an exchange economy with continuous and locally non-satiated preferences and time discounting.

insights into implementation in dynamic settings, it leaves open many interesting questions associated with the problems of dynamic implementation.

References

[1] Arya, A., J. Glover, and J. Hughes (1997), "Implementing Coordinated Team Play," *Journal of Economic Theory*, 74, pp. 218–232.

[2] Baliga, S., L. Corchon, and T. Sjöström (1995), "The Theory of Implementation When the Planner is a Player," forthcoming in *Journal of Economic Theory.*

[3] Baliga, S. and T. Sjöström (1995), "Interactive Implementation," mimeo: Harvard University

[4] Baron, D. and J. Ferejohn (1989) "Bargaining in Legislatures," *American Political Science Review*, vol. 83, pp. 1181-1206.

[5] Chakravorti, B., L. Corchon, and S. Wilkie (1992), "Credible Implementation," forthcoming in *Games and Economic Behavior.*

[6] Fudenberg, D. and J. Tirole (1991), *Game Theory*, MIT Press: Cambridge, MA.

[7] Hurwicz, L. (1972), "On Informationally Decentralized Systems," In C. B. McGuire and R. Radner, *Decision and Organization*, North Holland: Amsterdam.

[8] Hurwicz, L. (1994), "Economic Design, Adjustment Processes, Mechanisms and Institutions," *Economic Design*, 1, pp. 1–14.

[9] Hurwicz, L., E. Maskin, and A. Postlewaite (1995), "Feasible Nash Implementation of Social Choice Correspondences when the Designer Does Not Know Endowments or Production Sets," In J. Ledyard, *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability*, Kluwer Academic Publishers: Amsterdam.

[10] Jackson, M. O. and T. R. Palfrey (1998), "Efficiency and Voluntary Implementation in Markets with Repeated Pairwise Bargaining," *Econometrica*, 66, pp. 1353–1388.

[11] Jackson, M. O. (1997), "A Crash Course in Implementation Theory," mimeo: California Institute of Technology.

[12] Kalai, E. and J. Ledyard (1998), "Repeated Implementation," *Journal of Economic Theory*, 83, pp. 308–317.

[13] Ma, C., J. Moore, and S. Turnbull (1988), "Stopping Agents from 'Cheating," *Journal of Economic Theory,* 46, pp. 355–72.

[14] Maskin, E. (1998), "Nash Equilibrium and Welfare Optimality", *Review of Economics Studies*, 66, pp. 23–38.

[15] Maskin, E. and J. Moore (1998), "Implementation with Renegotiation," *Review of Economics Studies*, 66, pp. 39–56.

[16] Moore, J. and R. Repullo (1988), "Subgame Perfect Implementation," *Econometrica*, 56, pp. 1191–1220.

[17] Rubinstein, A. and A. Wolinsky (1992), "Renegotiation-Proof Implementation and Time Preferences," *American Economic Review.*

Appendix

Proof of Theorem 2 The message spaces are defined by:

$$M_i = S \times A \times \{0, 1, 2, \ldots\}$$

The outcome function is defined by partitioning the set of message profiles into two regions corresponding to compatible (D_1) and incompatible (D_2) messages.

$$D_1 = \{m \in M | \exists i \in I, \;\; s \in S, \;\; z \in A \text{ such that } m_j = (s, z, 0) \text{ for all } j \neq i,$$

$$\text{and } G(z, s) \in F(s)\}$$

$$D_2 = \{m \in M | m \notin D_1\}$$

In D_1, there can be at most one deviator. If there is no deviator, then $g(m) = z$. If i is the single deviator, denote $m_i = (s', y, k)$ and define the outcome function as:

$$\begin{aligned} g(m) &= z \text{ if } G(y, s) P_i(s) G(z, s) \text{ and } m \in D_1 \\ &= y \text{ if } G(z, s) R_i(s) G(y, s) \text{ and } m \in D_1 \end{aligned}$$

In D_2, the outcome is determined by the largest integer game:

$$g(m) = m_{i^*}^2$$

where $i^* = \min\{i | m_i^3 \geq m_j^j \text{ for all } j \in I\}$. The proof now involves showing three things:

1. **If $a \in F(s)$ then there is a Nash equilibrium at s in which** $m_i = (s, a, 0)$ for all i. To see this, first observe that since $a \in F(s)$, it must be that $a = G(a, s)$, so $m \in D_1$, $g(m) = a$, and $G(g(m), s) = a$. Furthermore, since s is being reported truthfully, any unilateral deviation by some player i to $\widehat{m}_i = (s', y, k)$ can only change the outcome if $G(y, s) \neq a$ and $aR_i(s)G(y, s)$. In this case, i is no better off than he would have been reporting $(s, a, 0)$. Thus, it is an equilibrium for everyone to report $(s, a, 0)$.

2. **If $G(a, s) \notin F(s')$ then $m_i = (s', a, 0)$ for all i is not a Nash equilibrium at s'.** This follows from h-monotonicity, which guarantees the existence of a feasible outcome,

y, and an individual, i, such that $aR_i(s)G(y,s)$ and $G(y,s',h)P_i(s')G(a,s',h)$. By doing making a unilateral deviation to (s',y,k), i can change the outcome from $G(a,s',h)$ to $G(y,s',h)$, which makes i strictly better off.

3. **If m^* is an equilibrium at s and it is not the case that $m_i = (s,z,0)$ for all i and $G(z,s) \in F(s)$ then $G(g(m^*),s) \in F(s)$.** Since it is not the case that $m_i = (s,z,0)$ for all i and $G(z,s) \in F(s)$, then at least $n \Leftrightarrow 1$ of the agents can unilaterally cause the outcome function to choose any outcome in A. Thus, for m^* to be an equilibrium requires that $G(g(m^*),s)R_j(s)G(y,s)$ for all $y \in A$, which, by $h \Leftrightarrow NVP$, implies that $G(g(m^*),s) \in F(s)$.

∎

Relationship of Stationary Equilibrium to Markov Perfect Equilibrium

Given a mechanism $[M,g]$ define the dynamic (stochastic) version of the mechanism, $[M,g]^\infty$, as follows. We append n veto moves (one for each player) to $[M,g]$, producing an $n+1$ stage game form. This is then played and terminates if no player vetoes and is repeated otherwise. That is, in stage 1, each player i independently submits a message $m_i^1 \in M_i$. After this is done m^1 is revealed to all players. Next, player 1 chooses $v_1^1 \in \{0,1\}$. The other players observe v_1^1 and then player 2 chooses $v_2^1 \in \{0,1\}$. The other players observe this and then player 3 chooses $v_3^1 \in \{0,1\}$. This process continues until all players have made a veto choice v_i^1. If $v_i^1 = 0$ for all i then the game form ends and the outcome is $g(m)^1$. Otherwise, play proceeds to period 2 and the process starts over: players again report messages followed by a sequence of veto moves. If in the second stage $v_i^2 = 0$ for all i then the game form ends and the outcome is $g(m)^2$. If not, then play proceeds to period 3, and so on. Thus, the interpretation is that $v_i^t = 1$ constitutes a veto by i of $g(m^t)$.

Fixing any given preference profile, R, $[M,g]^\infty$ and R define a stochastic game, and so standard definitions of pure strategies and subgame perfect equilibrium in pure strategies apply.[16] For this game, a pure Markov strategy for i is a choice of m_i, and $v_i : M \times \{0,1\}^{i-1} \rightarrow$

[16]See Fudenberg and Tirole (1991) for definitions. The stochastic nature of the game here is a very simple

$\{0,1\}$ that is to played in every period (such that the game has not already ended).[17] Thus, $v_i(m, v_1, \ldots, v_{i-1})$ is a function of the message profile and the veto decisions of the previous players in the current period. This can be further simplified, as the only time a player's veto makes a difference is in the case where the previous veto choices where all 0. Thus, without loss of generality for the definition of Markov strategy in this game, one can take v_i to depend only on m. The corresponding strategy in the infinite game is found by simply having i play m_i and v_i whenever called on to do so. A pure strategy Markov perfect equilibrium $\{R, [M,g]^\infty\}$ is a profile of pure Markov strategies which form a pure strategy subgame perfect equilibrium of $\{R, [M,g]^\infty\}$.

Lemma 1 *If m is a stationary equilibrium of $[M,g]$ at s, then m together with v, defined by $v_i(\overline{m}) = 1$ if and only if $g(m)^{t+1}P_i(s)g(\overline{m})^t$ for each i, form a Markov perfect equilibrium of $\{R(s), [M,g]^\infty\}$. Conversely, if (m, v) is a pure strategy Markov perfect equilibrium of $\{R(s), [M,g]^\infty\}$ and v is such that $v_i(\overline{m}) = 0$ if $g(\overline{m})^t I^i(s) g(m)^{t+1}$, then m is a stationary equilibrium of $[M,g]$ at s.*

Note that if $v_i(\overline{m}) = 0$ if $g(\overline{m})^t I^i(s) g(m)^{t+1}$, then it must be that $v_i(\overline{m}) = 0$ if and only if $g(\overline{m})^t R^i(s) g(m)^{t+1}$. So the converse requires that players not veto when they are indifferent.

To see why this is necessary consider the following simple example. $M_1 = \{m_1, \overline{m}_1\}$ and $M_2 = \{m_2\}$. Let $g(m_1, m_2) = a$ and $g(\overline{m}_1, m_2) = b$. Preferences are such that $a^t P_1(s) b^t P_1(s) b^{t+1} P_1(s) \emptyset$ and $a^t I_2(s) b^t I_2(s) b^{t+1} P_2(s) \emptyset$. So player 2 is completely indifferent while player 1 prefers a to b. The only stationary equilibrium is m_1, m_2. The combination $\overline{m}_1, m_2$ is not a stationary as 1 has an improving deviation to m_1, m_2 which will not be vetoed since 1 prefers a^t to b^{t+1} and 2 is indifferent. However, there is a Markov perfect equilibrium with $\overline{m}_1, m_2$ where player 2 always vetoes m_1, m_2, since player 2 is fully indifferent.

one with two states {ended, continue}, which keep track of whether there has been a period with no vetoes and so the game has ended, or whether in each previous period someone has vetoed and the game is continuing. We consider only pure strategies and only pure strategy deviations.

[17] Again, see Fudenberg and Tirole (1991) for discussion of Markov strategies, Markov perfect equilibrium, and references.

Proof: Let us first show that if m is a stationary equilibrium of $[M, g]$ at s, then m together with v, defined by $v_i(\overline{m}) = 1$ if and only if $g(m)^{t+1} P_i(s) g(\overline{m})^t$ for each i, form a Markov perfect equilibrium of $\{R(s), [M, g]^\infty\}$. First, it follows directly from the definition of stationary equilibrium that if there is an improving deviation from the Markov strategies m, v for some player i, then that deviation must involve a deviation in more than one period. The finite one stage deviation principle (see, e.g., the proof of theorem 4.1 in Fudenberg and Tirole (1991)) then implies that if there is an improving deviation it must result in the outcome ∞. However, $g(m)$ is (weakly) preferred to ∞ and so m, v is a Markov perfect equilibrium of $\{R(s), [M, g]^\infty\}$.

Let us now show that if m, v is a pure strategy Markov perfect equilibrium of $\{R(s), [M, g]^\infty\}$ and v is such that $v_i(\overline{m}) = 0$ if $g(\overline{m})^t I_i(s) g(m)^{t+1}$, then m is a stationary equilibrium of $[M, g]$ at s. First note that if (m, v) is a pure strategy equilibrium of $\{R(s), [M, g]^\infty\}$ then $v_i(m) = 0$ for each i and so the outcome is $g(m)$ at time 1. (The only alternative outcome is $\emptyset$, and since all individuals prefer $g(m)$ to $\emptyset$, subgame perfection implies that each player conditional on no previous vetoes must not veto $g(m)$.) Also, subgame perfection and the fact that $v_j(\overline{m}) = 0$ if $g(\overline{m})^t I_j(s) g(m)^{t+1}$, implies that $v_j(\overline{m}) = 0$ if $g(\overline{m})^t R_j(s) g(m)^{t+1}$. Suppose that m is not a stationary equilibrium of $[M, g]$ at s. Then there exists i such that $H(g(\widehat{m}_i, m_{-i}), s, g(m)_2) P_i(s) g(m)$. This implies that $H(g(\widehat{m}_i, m_{-i}), s, g(m)_2) = g(\widehat{m}_i, m_{-i})$ and so $g(\widehat{m}_i, m_{-i}) P_i(s) g(m)$ and $g(\widehat{m}_i, m_{-i}) R_j(s) g(m)_2$ for all j. Since $v_j(\overline{m}) = 0$ if $g(\overline{m})^t R_j(s) g(m)^{t+1}$ for each j, we know that $v_j(\widehat{m}_i, m_{-i}) = 0$ for each j. But then $g(\widehat{m}_i, m_{-i}) P_i(s) g(m)$ contradicts the fact that m, v form a Markov perfect equilibrium. ∎